Documentation of data

From metadata to tools

Øystein Godøy, Markus Fiebig, Torill Hamre

Topics

- Metadata
  - Discovery level
  - Use level
- Data structure and formatting
- Provenance
- ***Tools***
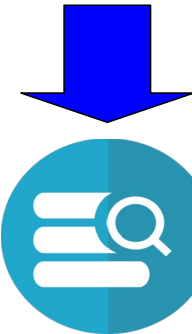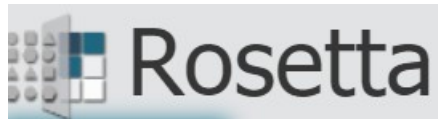
Tools

- Why tools?
  - To add documentation to dataset files
    - Discovery metadata
    - Usage metadata
  - To prepare datasets in standard formats
  - Makes data easier to reuse (also by others)

- What tools?
  - GIS (commercial, free)
  - Data processing and analysis tools (commercial, free)
  - Rosetta (free)

Tools

- Rosetta
  - Open source tool, written in Java
    - https://github.com/Unidata/rosetta
  - Developed by UniData
  - Web-based application
  - Converts ASCII files to NetCDF
  - Can read metadata from header(s)
  - Can add user defined metadata
  - Saves "setup" as templates for future

Rosetta

Rosetta is Beta software under active development, use at your own risk. This specific version of Rosetta has been tailored for NMDC.

Welcome to Rosetta, a data transformation tool. Rosetta is a web-based service that provides an easy, wizard-based interface for data collectors to transform their datalogger generated ASCII output into Climate and Forecast (CF) compliant netCDF files. These files will contain the metadata describing what data is contained in the file, the instruments used to collect the data, and other critical information that otherwise may be lost in one of many dreaded README files.

In addition, with the understanding that the observational community does appreciate the ease of use of ASCII files, methods for transforming the netCDF back into a user defined CSV or spreadsheet formats is planned to be incorporated into Rosetta.

We hope that Rosetta will be of value to the science community users who have needs for transforming the data they have collected or stored in non-standard formats.

Rosetta is currently under continued further development, and ready for beta testing.

What would you like to do?

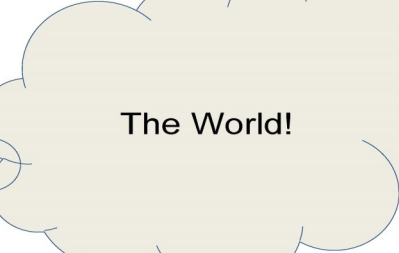[Convert a file to the netCDF format and create a new template]
[Upload, modify, and use an existing template]

Version : 0.5.1-NERSC
Build Date: Fri Aug 03 11:25:13 CEST 2018

UCAR, Unidata Program Center

Arms et al., 2015. *Rosetta: Ensuring the Preservation and usability of ASCII-based Data into the Future.* Poster at AGU 2015.

The World!

Publish to a data server

Download Converted File
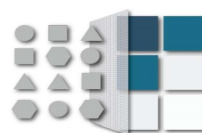
Rosetta

Transformation Workflow

Add General Metadata

Choose data collection method

Select the data header

Define Data Variables and Metadata

Define Coordinate Variables and Metadata

Tools

- Example
  - Sailbuoy data 2016
  - Converted from ASCII to NetCDF with Rosetta
  - Published in NMDC through Thredds Data Server (TDS)
  - Searchable through NMDC Data Portal at https://nmdc.no/

Ocean Acidification data in 2016 in the Fram Strait (RFF Iskantseilas)

- Collected by a remotely controlled sailbuoy
- Parameters:
  - temperature,
  - conductivity,
  - pH
  - partial pressure of carbon dioxide (pCO2),
  - dissolved oxygen (O2).
- Every 10 minutes
- 30 June - 18 July 2016
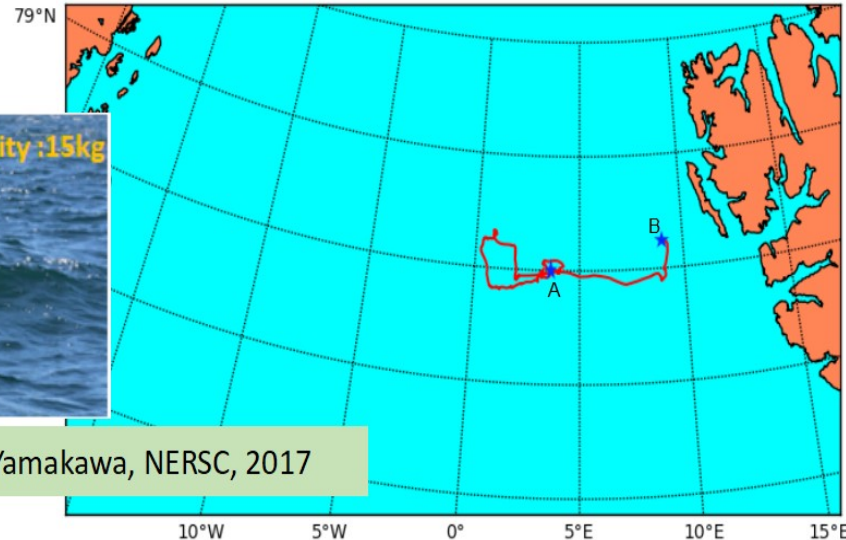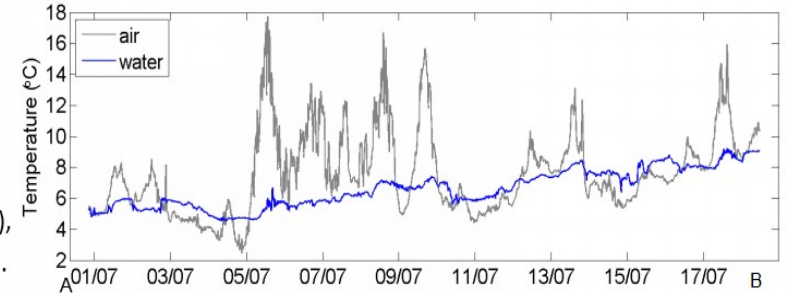
Courtesy of J. Ullgren and A. Yamakawa, NERSC, 2017

NorDataNet
Norwegian Scientific Data Network
GeoAccessNO

Tools

- What should motivate a tool developer?

  - To serve the needs of the users!

  - Be like a good football referee (i.e. "invisible")

  - Because, how much fun really is format conversion?

  - Luckily we can use tools like Rosetta

  - Most of us would rather do something else...

E   F

COFFEE
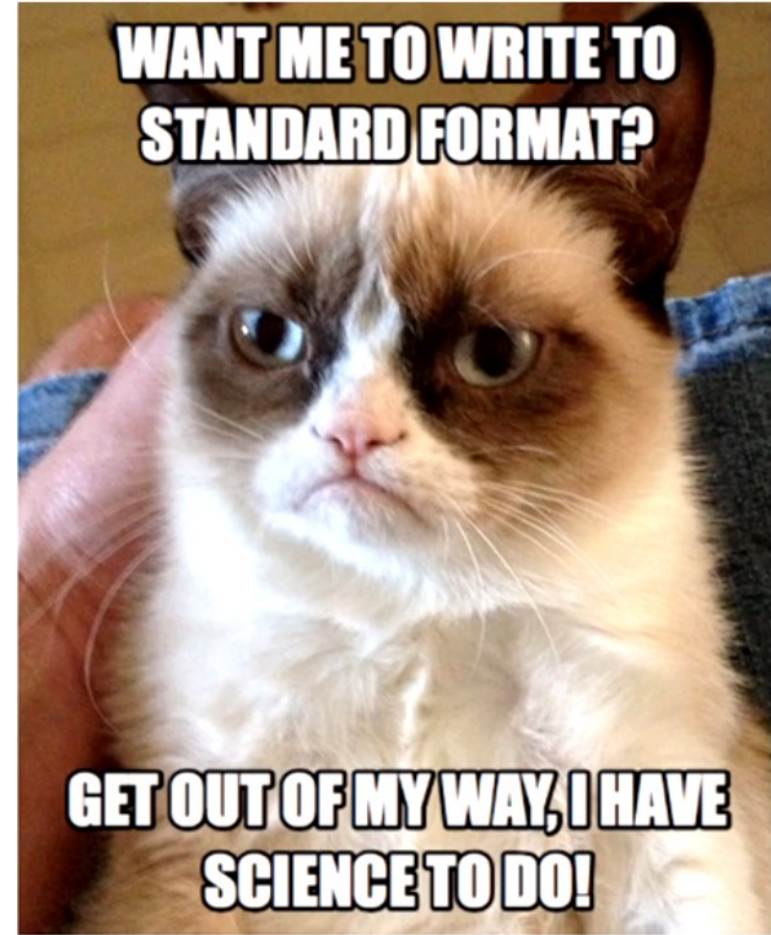
STOP HERE
TO REFUEL
(TAKE OUT AVAILABLE)